**Nitte Meenakshi Institute of Technology**

Department of Computer Science and Engineering

**18CSE751 Introduction to Machine Learning Learning Activity Proposal**

**Movie Recommendation System using KNN**

**Anish Pokhrel [1NT118CS191]**

**Shreeyut Shrestha [1NT18CS206]**

**TejNarayan Chauhan [1NT18CS209]**

**Abstract**

Recommendation systems are becoming increasingly important in today’s hectic world. People are always in the lookout for products/services that are best suited for them. Therefore, the recommendation systems are important as they help them make the right choices, without having to expend their cognitive resources.In this system recommendation system we will predict the rating of the given movie based on its neighbors and compare it with the actual rating.

# Introduction

A recommender system, or a recommendation system is a subclass of information filtering system that seeks to predict the "rating" or "preference" a user would give to an item. Recommender systems are used in a variety of areas, with commonly recognized examples taking the form of playlist generators for video and music services, product recommenders for online stores, or content recommenders for social media platforms and open web content recommenders. These systems can operate using a single input, like music, or multiple inputs within and across platforms like news, books, and search queries. There are also popular recommender systems for specific topics like restaurants and online dating. Recommender systems have also been developed to explore research articles and experts, collaborators, and financial services.

# Data Set

We will use the dataset from the Donate-a-cry campaign. The dataset is obtained from the following link –

<https://grouplens.org/datasets/movielens/latest/>

This dataset (ml-latest-small) describes 5-star rating and free-text tagging activity from [MovieLens](http://movielens.org/), a movie recommendation service. It contains 100836 ratings and 3683 tag applications across 9742 movies. These data were created by 610 users between March 29, 1996 and September 24, 2018. This dataset was generated on September 26, 2018.

The content in the file are as follows

1. User ID : MovieLens users were selected at random for inclusion. Their ids have been anonymized. User ids are consistent between ratings.csv and tags.csv .
2. Movie ID : Only movies with at least one rating or tag are included in the dataset. These movie ids are consistent with those used on the MovieLens web site.
3. Ratings Data File Structure : All ratings are contained in the file ratings.csv. Each line of this file after the header row represents one rating of one movie by one user, and has the following format:

userId,movieId,rating,timestamp

1. Movie Data File Structure : Movie information is contained in the file movies.csv. Each line of this file after the header row represents one movie, and has the following format:

movieId,title,genres

Genres are a pipe-separated list, and are selected from the following:

* Action
* Adventure
* Animation
* Children's
* Comedy
* Crime
* Documentary
* Drama
* Fantasy
* Film-Noir
* Horror
* Musical
* Mystery
* Romance
* Sci-Fi
* Thriller
* War
* Western
* (no genres listed)

1. Links Data File Structure : Identifiers that can be used to link to other sources of movie data are contained in the file links.csv. Each line of this file after the header row represents one movie, and has the following format:

movieId,imdbId,tmdbId

# Machine Learning Methods

* **KNN:** K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique. K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most like the available categories. K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm. K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems. K-NN is a non-parametric algorithm, which means it does not make any assumption on underlying data. It is also called a lazy learner algorithm because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset. KNN algorithm at the training

phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much like the new data.

# Assessment:

Rating scores of the model will be used for the assessment.

# Presentation and Visualization

Rating scores of the model will be displayed based on users input.

# Roles

Anish Pokhrel -: KNN description

Shreeyut Shrestha -: KNN implementation

TejNarayan Chauhan -:KNN working mechanism demonstration

# Schedule

The schedule is a table of dates and tasks that you plan to complete.

# Date Tasks to be Completed

17/01/21 Tasks completed by chosen date

18/01/22 Tasks to be completed by the final report/ presentation date

# Bibliography

* + <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8222524/>
  + [https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-](https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761) [algorithm-6a6e71d01761](https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761)
  + https://www.analyticsvidhya.com/blog/2020/08/recommendation-system-k-nearest-neighbors/